Maryland Population Research Center

WORKING PAPER

Addressing Abortion Underreporting in Surveys with the List Experiment: Lifetime and Five-Year Abortion Incidence with Multivariate Estimation of Socio-demographic and Health Associations in two U.S. States

PWP-MPRC-2022-001

November 2022

Authors:

Heide M. Jackson
Michael S. Rendall
University of Maryland

Addressing Abortion Underreporting in Surveys with the List Experiment: Lifetime and Five-Year Abortion Incidence with Multivariate Estimation of Socio-demographic and Health Associations in two U.S. States

Heide M. Jackson[1] and Michael S. Rendall[2]

1. Maryland Population Research Center, University of Maryland College Park
2. Department of Sociology and Maryland Population Research Center, University of Maryland College Park

Corresponding Author Email: heidej@umd.edu

Acknowledgements

This work was supported by infrastructural support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, population research infrastructure grant P2C-HD041041, and a research grant from an anonymous private philanthropic foundation. We thank Deonte Hughes for valuable research assistance.

Key words:

abortion, indirect survey methods, reproductive health

Abstract

Direct survey reports are known to produce severely downwardly biased estimates of women's abortion incidence. In the present study, incidence is estimated instead using an indirect survey method, the list experiment. The method is applied to cross-sectional surveys in two U.S. states to estimate cumulative-lifetime abortion incidence and incidence over the past five years. The five-year incidence estimates are evaluated against external benchmark data. Multivariate estimation of both five-year and cumulative-lifetime incidence, controlling for age and state of residence, is conducted for both standard socio-demographic characteristics and for health and health-care access predictors. The list-experiment estimates of five-year abortion incidence are found to be similar to the external benchmark data for both states. Expected age patterns of an inverse U-shape for five-year incidence, and a monotonic linear increase for cumulative lifetime incidence, are additionally produced. Controlling for these age relationships, expected patterns of disparities in abortion incidence by race and socio-demographic disadvantage are found for both five-year and cumulative lifetime incidence. Additionally, both lower self-rated health and difficulty obtaining either general health-care or birth control are predictive of higher abortion incidence. We conclude positively about the validity and utility of the list experiment method.

Introduction

Abortion is a critical aspect of reproductive health. From 2015-2019 approximately 34% of unintended pregnancies ended in abortion (R. Jones et al., 2022). In 2020, some 21% of all pregnancies ended in abortion (R. Jones et al., 2022). Although abortion is an intensely regulated, and in some states, now criminalized procedure (Cohen et al., 2022), relatively little is known about the characteristics of women most at risk of abortion (Ahrens & Hutcheon, 2020), in particular because direct question approaches in surveys yield large overall underestimates of abortion and likely differentials in underreporting by observable and unobservable characteristics. A recent analysis of three population-representative surveys found that direct questions used in the National Survey of Family Growth (NSFG), National Longitudinal Survey of Youth (NLSY) and National Survey of Adolescent to Adult Health (AddHealth) capture only 30-40% of abortions contained in external data (Lindberg et al., 2020). In the general NSFG survey, just 38% of pregnancies ending in abortion are captured relative to Guttmacher Abortion Provider Census estimates (Center for Disease Control, 2022). Because of this, the NSFG issues guidance to users not to use abortion data from the NSFG for substantive research projects (Center for Disease Control, 2022). Abortion underreporting is consistently found across U.S. states (Maddow-Zimet et al., 2021). Differential underreporting by observed socio-demographic characteristics has also been documented (Jagannathan, 2011; Lindberg et al., 2020; Udry et al., 1996) with underreporting on unobserved factors also likely (Lindberg et al., 2020). Simulations show that moderate amounts of differential underreporting may lead to bias in the estimated characteristics associated with abortion incidence in surveys (Lindberg et al., 2020). This failure of direct questions to adequately measure abortion has been found in several decades of survey research (Jagannathan, 2011; E. F. Jones & Forrest, 1992; R. K. Jones & Kost, 2007; Lindberg et al., 2020; Tierney, 2019). As a consequence, population representative survey data are seldom used to study substantive questions related to abortion. Indeed we are only aware of two studies from recent decades (Gius, 2007; Sutton et al., 2019) that use survey data for studying abortion, with the latter performing extensive additional analyses from alternate data sources to validate their estimation using the NLSY's direct question.

Much of what is known about levels and socio-demographic characteristics comes from using Guttmacher Institute censuses and surveys: the Abortion Provider Census (APC) capturing overall abortion counts by state, and the periodic Abortion Patient Survey (APS) that surveys a nationally representative sample of abortion patients. These data have together been used to estimate overall trends in annual abortion incidence by state, lifetime abortion incidence for synthetic cohorts, and annual abortion incidence by characteristics of abortion patients. For the latter, APS abortion numerators are matched to at-risk population denominators from the American Community Survey (ACS) and NSFG (R. K. Jones & Jerman, 2017), and from the Current Population Survey (CPS) and NSFG (Jones and Kavanaugh 2011). As valuable as are these APS-based estimates, they are nevertheless limited to characteristics for which there is a suitable denominator definition in large-scale survey data, and they do not readily allow for multivariate estimation.

In the present study, we introduce an indirect survey approach, the abortion list experiment, and apply it using population-representative surveys of reproductive age (18-44) women living in two states, Delaware and Maryland, in 2017 and in 2021. The list experiment for these surveys respectively measures cumulative lifetime incidence in 2017 and incidence over the past five years in 2021. We evaluate the validity of the five-year list-experiment estimates of overall abortion incidence against estimates that we calculate for the same period for

Delaware and Maryland using external data sources. We compare age-patterns of five-year abortion incidence estimated using the list experiment to annual-incidence age patterns in CDC and APS-based data, and we compare list-experiment estimates of abortion incidence by socio-demographic characteristics in Delaware and Maryland to national estimates of abortion differences for these socio-demographic characteristics using APS-based estimation. These include variables that are amenable to constructing population denominators from large-scale survey sources such as race, education, income, marital status, and parity. A novel contribution of our list-experiment estimation is that we are additionally able to estimate abortion differences for variables that typically are not amenable to constructing population denominators from other sources. We demonstrate this for self-rated health and experience of difficulties in accessing health care. Other novel contributions of our study include the evaluation of the statistical feasibility of the abortion list-experiment estimator over a shorter, five-year period. Previous work using list-experiment estimation has estimated only cumulative lifetime abortion incidence. Cumulative lifetime incidence may not be suited to studying time-varying predictors of abortion incidence. We also show the feasibility of combining five-year and cumulative-lifetime estimates to calculate lifetime abortion incidence of real cohorts attaining age 40-44 in Delaware and Maryland in 2021.

Our study also introduces multivariate regression modeling to U.S. abortion list estimation, including controls for age and state of residence when estimating the above associations of socio-demographic, health and health-care access characteristics. Previously, multivariate regression modeling of abortion list estimation has been in non-U.S. settings (Moseson et al., 2017). Multivariate regression applications to list experiment estimation on topics other than abortion, in the U.S. and elsewhere, have had mixed success (Ahlquist, 2018; Wolter & Laier, 2014).

Literature Review

Survey data are crucial for their use in understanding differences in risk of abortion and mechanisms associated with these risks (Ahrens & Hutcheon, 2020; Dehlendorf et al., 2013). With survey data, abortion can be contextualized with the characteristics, attitudes, and behaviors of individuals who either do or do not have an abortion over a given interval of interest. However, surveys have been shown to consistently under-estimate abortion incidence (E. F. Jones & Forrest, 1992; Lindberg et al., 2020; Udry et al., 1996). Additionally, some studies have found evidence of differential underreporting of abortions (Desai et al., 2021; Jagannathan, 2011) such that the abortion histories of unmarried, Black, and low income women are particularly under-estimated. Moreover, the underreporting of abortions may lead to differential under-estimates of all pregnancies and the timing between pregnancies (Desai et al., 2021).

For standard socio-demographic characteristics, researchers using the APS have been able to construct population denominators from external sources (R. K. Jones & Jerman, 2017; R. Jones & Kavanaugh, 2011), or have been able to use other dimensions of the APS data to study disparities, for example, comparing abortions by using gestation length as an indicator of disparities in access (Solazzo, 2019). Health-care access variables such as insurance coverage and distance lived from the abortion clinic have also been used as predictors of self-reported prior abortions for those women presenting at a clinic for a current abortion (R. Jones et al., 2018).

For many characteristics and experiences potentially relevant to differentiating abortion incidence, however, external population denominators may not be available, and alternative analytical strategies cannot be readily adapted. For example, discrimination in healthcare access has been proposed as a mechanism leading to higher abortion incidence (Dehlendorf et al., 2013), but the size of the population experiencing discrimination in health care access is not readily obtained from external data sources.

Linking APS numerators with population denominators calculated from other surveys such as the ACS, CPS, or NSFG also does not readily allow for multivariate estimation. As such, these methods do not allow inferences about theoretically relevant variables adjusting for other demographic characteristics. For example, it is often found that women with no prior births (“nulliparous) have a lower abortion incidence than their parous contemporaries (R. K. Jones & Jerman, 2017; R. Jones & Kavanaugh, 2011); however, nulliparous women are also, on average, younger. Multivariate methods are needed to disentangle whether the association between parity and abortion incidence remains after accounting for age differences.

Analysis of population-representative survey data such as the NSFG, NLSY, and Add Health could overcome both limitations of APS-based methods for understanding abortion differentials. However, because of the severe bias found with direct survey questions, researchers have been motivated to consider indirect methods of collecting abortion data in population-representative surveys.

Indirect Survey Methods for Abortion Estimation: The List Experiment

Indirect approaches for measurement in surveys have a long history in estimating a range of sensitive behavioral and attitudinal topics (Tourangeau & Yan, 2007). The list experiment has been developed for sensitive topics such as the self-reporting of criminal and delinquent behavior and for political attitudes and voting behavior (Blair et al., 2020; Ehler et al., 2021). For abortion, indirect approaches including the list experiment, the best friend approach, confidante method, network scale up method, and the randomized response technique have been implemented in developing country settings (Lara et al., 2004; Moseson et al., 2015; Stillman et al., 2020; Sully et al., 2020; Yeatman & Trinitapoli, 2011). Each have their limitations, however, and accordingly have received limited attention in the U.S. context.

In the list experiment, also known as the unmatched count technique or the item count technique (Ehler et al., 2021), respondents are randomly assigned to receive a treatment or control list and are asked how many, but not which, items on that list apply to them. In abortion list experiments, respondents receiving the control list are asked about a series of non-stigmatized health-related ('non-sensitive') items, and respondents receiving the treatment list are asked about the same non-sensitive health items and the sensitive item of whether they have had an abortion. The difference in mean number of items reported between treatment and control lists captures the incidence of the sensitive item, having had an abortion. In order to maximize statistical power (Blair et al., 2020), double list experiments are often used. In the double list experiment, each respondent receives a treatment list and a distinct control list (for an example, see Table 1). In order for list experiments to generate valid inferences, respondents must be honest in their reporting of the sensitive item and their reporting of non-sensitive items must not be affected by the inclusion of the sensitive item on the list (Blair & Imai, 2012). Further, in order to avoid sensitive item disclosure, the list should be formed such that few respondents have none or all of the items apply to them (Kuklinski et al., 1997).

Compared with other indirect estimation methods such as the best friend approach (Yeatman & Trinitapoli, 2011), network scale up method (Sully et al., 2020), and the randomized response technique (Lara et al., 2004), list experiments are relatively inexpensive to implement and easy to analyze; however, they have two notable limitations. First, the list experiment approach does not provide information on whether a specific respondent had an abortion. As such the list experiment has limited utility for studies examining health or social outcomes by whether a woman has had an abortion ----- that is, using abortion as a predictor variable ---- although methods have been developed to partially overcome this limitation (Imai et al., 2015). Second, the list experiment is statistically inefficient relative to direct question approaches. Blair et al. (2020) explicate this efficiency loss as follows:

For a direct question with unobserved incidence τ^* and sensitivity bias, represented by $bias(\tau^*, \tau)$, the sample variance with a direct question is

$$V(\hat{\tau}) = \frac{1}{N-1}\{\tau^*(1-\tau^*) + bias(\tau^*, \tau)(1 - bias(\tau^*, \tau)) - 2(bias(\tau^*, \tau) - \tau^* bias(\tau^*, \tau))\} \quad (1)$$

For the list experiment, the variance is

$$V(\widehat{\tau^*}) = \frac{1}{N-1}\{\tau^*(1-\tau^*) + 4V(Y_i(0)\big) + 4Cov(Y_i(0), D_i^*) \quad (2)$$

where Y_i is the response to the non-sensitive items on the list, $D_i^*$is the unobserved sensitive item, V(Y_i(0)) is the variance between the non-sensitive items, and CovY_i(0) D_i^*) is the covariance between the non-sensitive and sensitive items. The notation * is used to denote those items which are not directly observed. The variance of the list experiment becomes higher when the incidence of the sensitive item is especially low (Droitcour et al., 1991) or the correlation between non-sensitive items is positive. While the inclusion of a double list experiment reduces variance by half (Blair et al., 2020), the increase in variance remains substantial relative to the direct question approach. This problem is further exacerbated for any sub-group analyses that further stratify the survey sample.

To our knowledge, only two studies have used list experiments to measure abortion incidence in the U.S. using population-representative data (Hood et al., 2022; Kissling & Jackson, 2022). Both studies use the list experiment to estimate cumulative lifetime abortion incidence and neither use multivariate estimation. We build on this previous work by estimating not only cumulative lifetime abortion incidence but also incidence in the five years prior to the survey. Moreover, we show how these two types of estimates can be combined for estimation of lifetime abortion incidence in real cohorts. Additionally, we use multivariate estimation to improve statistical efficiency and to account for confounding predictors. Until now, multivariate regression has rarely been applied in abortion studies (Moseson et al., 2017) and has been met with mixed success in applications to other topic areas (Ahlquist, 2018; Wolter & Laier, 2014).

Data

We use data from the Statewide Survey of Women of Reproductive Age in Delaware and Maryland (SoW) (Boudreaux et al., 2022; Boudreaux & Rendall, 2020). The SoW is a repeated cross-sectional survey of women aged 18 to 44. The survey mode was a combination of internet-completed and mailed self-administered questionnaires (about three quarters internet-completed and one quarter mailed), plus a very small number of telephone interviews. It was administered to stratified random samples of 1,496 women in Delaware and 1,451 in Maryland from November 2016 through March 2017 ("Baseline Survey", NORC 2019) and to a second set of stratified random samples of 4,063 women in Delaware and 3,078 women in Maryland from February 2021 through September 2021 ("Endline Survey", NORC 2021). Minority race/ethnicity populations were oversampled in the stratification design. Overall response rates were 23.1% at Baseline and 26.7% at Endline. Post-stratification sample weights calibrated to Census Public Use Microdata allow SoW estimates to represent Delaware and Maryland women between the ages of 18 and 44 in 2017 and 2021 (NORC, 2019; NORC, 2021). In our analyses, we normalize these weights to 1 for each state before pooling observations across Delaware and Maryland.

[TABLE 1 ABOUT HERE]

The survey content includes standard questions on contraception, fertility history, sexual activity, and socio-demographic characteristics. Both Baseline and Endline SoWs, however, additionally included abortion double list experiment questions. Approximately one half of respondents were assigned to receive Treatment List A and Control List B and half were assigned to receive Control List A and Treatment List List B. Table 1 shows the list experiment items (see Moseson et al. (2019) for a previous usage of non-sensitive items). Lists included two high-incidence non-sensitive items and one low-incidence item to minimize the chances that respondents would have zero or all items apply to them, thus revealing whether they had or had not had an abortion (Kuklinski et al., 1997). Respondents in each treatment list group received an additional 'sensitive' item, which asked if they had had an abortion. At the Baseline survey, this additional item was: Ever had an abortion (ended a pregnancy on purpose); at the Endline survey, this item was: Had an abortion (ended a pregnancy on purpose) in the past 5 years.

Methods

We use the list experiment to estimate 2017 cumulative-lifetime abortion incidence and 2021 five-year abortion incidence by state and by state and age. We first estimate abortion incidence using a simple difference-in-means estimator. The difference-in-means estimator involves estimating abortion incidence (τ) as the difference between treatment and control lists. It is specified as:

$$\tau = \frac{1}{N_1}\sum_{i=1}^{N} T_i\, Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - T_i\,)\, Y_i \quad (3)$$

where $N_1 = \sum_{i=1}^{N} T_i$ is the size of the treatment group and N_0 is the size of the control group and $T_i Y_i$ captures treatment group responses with $(1 - T_i)Y_i$ capturing control group responses. We estimate cumulative-lifetime and five-year abortion incidence for the overall (all-ages) sample and for individual age groups (18-24, 25-29, 30-34, 35-39, 40-44) in each of the two states. Computationally, the difference-in-means calculation is equivalent to a linear regression that is estimated separately for a given age group and state of residence. Following prior studies (Bell & Bishai, 2019; Cowan et al., 2016; Kissling & Jackson, 2022) and to improve statistical power, we average estimates across the two lists.

In a second estimation approach, we increase statistical power by pooling data across states and use a multivariate regression with a linear estimator (Tsai, 2019). To estimate overall cumulative-lifetime incidence to age x (where x is a woman's age in 2017), our regression takes the form:

$$P(abortion_x) = \beta MD \quad (4)$$

To estimate five-year abortion incidence to 2021, that is abortions between age x in 2017 and age x+5 in 2021, our regression is:

$$P(\,_5abortion_x) = \beta MD \quad (5)$$

To estimate abortion incidence by both age and state, we specify additional models that parameterize age. For modeling cumulative lifetime abortion incidence, we expect a monotonically increasing relationship of age to cumulative lifetime abortion incidence. Formally, we model cumulative abortion incidence by age in 2017 with:

$$P(abortion_x) = \beta age + \beta MD \quad (6)$$

For modeling five-year abortion incidence, previous work (using APS data and state health department reports compiled by the CDC) finds that annual abortion incidence rises and then falls with age (R. K. Jones & Jerman, 2017; Kortsmit, 2021), motivating our quadratic age specification:

$$P({}_5abortion_x) = \beta age + \beta age^2 + \beta MD \quad (7)$$

We include state of residence in each model as an indicator variable, thereby imposing the same age relationship in each of the two states (except for a state-of-residence level shift). We use a linear estimator because, although it is less efficient than a maximum likelihood estimator, it has reduced bias when incidence is low (Ahlquist, 2018; Blair et al., 2019). Low incidence is especially the case when we estimate five-year abortion incidence.

Our specification of 'beta' parameters here is for exposition purposes. Our regression estimation for double list experiment data uses the *kict* package with a linear estimator in Stata (Tsai, 2019). Tsai's estimation adapts Imai's (2011) least squares estimation to accommodate a dual list experiment design. Coefficients are produced in a three-stage procedure. In the first two stages, two sets of gamma coefficients, one for each control list of non-sensitive items, are produced. In the third stage, these gamma coefficients are averaged in the estimation of the delta coefficients, which represent the change in incidence of the sensitive item (abortion) with a one unit change in the regressor. Because we are using a linear regression, the delta coefficient multiplied by 100 represent the percentage-point change in abortion incidence with a one unit increase in the predictor variable. The gamma coefficients which describe associations of the regressors with incidence of each of the non-sensitive items are also produced by the *kict* package, but we do not use these in our analyses.

Using coefficient values from the linear regression models, we predict abortion incidence overall and at the midpoint of the age intervals (i.e., ages 21, 27, 33, 37, 42). We evaluate the fit of these parametric age specifications both graphically and by using difference-in-means confidence intervals.

We additionally compare the overall (between the ages of 18-44) five-year incidence from our list experiment estimates to overall five-year incidence that we calculate from external data sources. Our external data calculations use abortion count numerators collected by the Guttmacher Abortion Provider Census (APC, (Maddow-Zimet & Kost, 2021) for total number of abortions by state by year, matched to population counts of women 18-44 living in these two states (obtained from the Census PUMS file American Community Survey for 2016-2020 (Ruggles, Steven et al., 2021)).

However, we also need to adjust the total abortions in 2016-2020 reported for Maryland and Delaware residents to reflect the number of abortions to distinct women, that is remove from the numerator of abortions women's repeat abortions in the interval. Specifically, we calculate:

$$2016 - 2020\ Abortion\ Incidence = \frac{Abortions\ to\ Distinct\ Women\ 18{-}44}{Mean\ 2016{-}2020\ Population\ of\ Women\ 18{-}44} \qquad (8)$$

APC data on total abortions had by Delaware and Maryland residents were directly collected by Guttmacher for calendar years 2016, 2017, 2019 and 2020. We estimate the number of 2018 abortions by averaging the number of abortions in 2017 and 2019. Our method for downwardly adjusting all abortions to abortions to distinct women is detailed in the supplementary appendix (Appendix 2). We show there that we estimate that 94.2% of abortions in Delaware and 94.0% of abortions in Maryland are to distinct women over the five-year period. We arrive at these estimated percentages by using distributions of abortions by order, from number of prior abortions reported by abortion patients collected by the Delaware state health department and published in CDC compilations (Kortsmit, 2021), and by using distributions of abortions by age from these same CDC compilations for Delaware and from imputed distributions for Maryland from the Guttmacher Institute (Maddow-Zimet & Kost, 2021), whose imputation averages over the age distributions of abortions in surrounding states.

The estimated percentages of total abortions that are to distinct women over the five-year period (94.2% and 94.0%) are relatively high in part because approximately 60% of all abortions in 2016-2020 are first abortions for the woman, leaving only the remaining 40% of abortions which may potentially be a higher-order abortion to occur during the five-year 2016-2020 period. Our method of estimating the probability of multiple abortions in the period relies on abortion incidence calculated at the mean age of abortion, where annual abortion rates are at their peak (see Appendix 2 and the Discussion section below). This reduces the likelihood of underestimation of repeat abortions and thereby reduces the likelihood of our overestimating abortions to distinct women.

Another goal of our study is to calculate lifetime abortion incidence for real cohorts. The Baseline SoW estimates cumulative lifetime abortion incidence at 2017 by age (and other characteristics) directly. We estimate for 2021 the cumulative lifetime abortion incidence at the midpoint of the 18-24, 25-29, 30-34, 35-39, and 40-44 age intervals by combining estimates from the Baseline and Endline SoWs for the same cohort. Our calculation uses the equation:

$$P(abortion_{x+5}) = P(abortion_x) + (1 - P(abortion_x)) * P({}_5abortion_x) \quad (9)$$

where $P(abortion_{x+5})$ is the abortion incidence in 2021, $P(abortion_{x})$ is the 2017 lifetime abortion incidence predicted using the coefficients from the linear regression model applied on the Baseline data, and $P({}_{5}abortion_{x})$ is the five-year abortion incidence predicted using the coefficients from the linear regression model on the Endline survey data. We generate 1000 bootstrapped samples and take the values at the 2.5 percentile and 97.5 percentile as our 95% confidence interval. To match the survey design, our bootstrap procedure stratifies by list assignment, state of residence, and survey stratum.

Finally, we extend our regression models to examine socio-demographic and health characteristics hypothesized to be associated with cumulative-lifetime and five-year abortion incidence. This is done using a separate regression model for each of the socio-demographic and health predictor variables, including controls for age and state of residence. We separately estimate regressions for race/ethnicity (Non-Hispanic Black, Non-Hispanic White, Hispanic or other race/ethnicity), household income (less than \$50,000, greater than or equal to \$50,000), education (less than a Bachelors, Bachelors degree or higher), partnership status (currently-married, currently-cohabiting, currently-single), parity (nulliparous or parous), self-rated health (Excellent, Very Good or Good versus Fair or Poor), and whether the woman experienced difficulties obtaining health care or in obtaining contraception in the last 12 months. The regressions control for age and state of residence using the parameterizations described in equations (6) and (7) above.

The validity of the list experiment method requires that several assumptions about randomization and response patterns be met (Blair et al., 2019). In Appendix 1, we describe these assumptions and results from statistical testing demonstrating that list experiment assumptions are met in the Endline (2021) survey (for testing done using the Baseline 2017 survey see Kissling & Jackson, 2022). We further demonstrate that the two sets of lists are mutually consistent in their estimates of abortion incidence (Lépine et al., 2020).

Results

We first present in Figure 1 annual abortion rates for Delaware, Maryland, and the U.S. from the Guttmacher Institute APC data from 2016-2020. Recent abortion rates are consistently higher in Maryland than in Delaware. While Delaware had a fairly unchanging abortion rate from 2016-2020 at about 17 abortions per 1000 women, the abortion rate in Maryland was consistently above 20 abortions per 1000 women and appeared to rise further in 2020. Both Maryland and Delaware have a higher abortion rate than is observed nationally. Abortion rates nationally had long been trending downward (Maddow Zimet & Kost, 2021), but rose slightly again between 2017 and 2020 (R. Jones et al, 2022).

[FIGURE 1 ABOUT HERE]

We first present our list-experiment estimates of abortion incidence in Delaware and Maryland overall and by age in each state. Figure 2 shows the cumulative lifetime abortion incidence up to 2017 separately by state. Presented for each state are (1) point estimates (circles) and confidence intervals from the difference-in-means estimator for each age group; and (2) regression-predicted point estimates (bars) from multivariate models (the regression coefficients and their standard errors are shown in Table 3). Overall (all ages 18-44) cumulative abortion incidence from our list experiment estimates are also shown. Figure 3 presents the equivalent estimates by age and overall for the five-year abortion incidence outcome. Figure 3 additionally includes overall five-year abortion incidence calculated from the external data sources (black diamonds).

[FIGURES 2 AND 3 ABOUT HERE]

Overall cumulative lifetime incidence to 2017, estimated from the difference-in-means approach, was 17.6% (CI: 10.9%-24.3%) in Delaware and 24.0% (CI: 18.6%-29.4%) in Maryland (Figure 2). Overall five-year abortion incidence to 2021 (Figure 3) was 5.8% in Delaware (CI: 2.5%-9.2%) and 10.1% in Maryland (CI: 5.9%-14.4%). These latter point estimates compare well to the five-year abortion incidence estimates from external data sources over the same period, of 8.1% for Delaware and 11.7% for Maryland. The survey confidence intervals for the five-year incidence calculated using the difference-in-means approach are seen to contain the values calculated from external data sources.

Age associations with abortion incidence are of the expected directions. Cumulative-lifetime abortion incidence increases monotonically with age, as can be seen in both the regression-predicted bars and in the overall pattern by age of the difference-in-means estimates (Figure 2). Five-year abortion incidence increases through ages 30-34 before subsequently falling, again seen in both difference-in-means estimates and the regression-predicted bars (Figure 3).

[FIGURE 4 ABOUT HERE]

We use the regression-predicted model estimates seen in Figures 2 and 3 also to produce cumulative lifetime abortion incidence estimates by age for real cohorts in 2017 and 2021. These are presented in Figure 4. The 2021 estimates combine the 2017 cumulative lifetime incidence to a given age group with the five-year incidence of the next age group. For example, the 35-39 year old cumulative lifetime incidence to 2017 is combined with five-year incidence for those aged 40-44 in 2021 to estimate lifetime incidence for the cohort attaining age 40-44 in 2021. In

2017, the Delaware lifetime abortion incidence was 28.2% (CI: 18.4%-37.1%) and the Maryland lifetime abortion incidence was 34.6% (CI: 26.7-41.8%). The 2021 lifetime abortion incidence estimates are similar to the respective 2017 lifetime incidences, at 27.1% (CI: 18.9%-34.7%) for Delaware and 36.3% (CI: 30.0%-42.6%) for Maryland. That is, just over a quarter of women in these Delaware cohorts, and just over a third of women in the equivalent Maryland cohorts, are estimated to have experienced an abortion in their lifetimes.

[TABLE 2 ABOUT HERE]

Table 2 shows distributions of women in Delaware and Maryland in 2017 and in 2021 by the characteristics for which we subsequently show estimates of abortion incidence. The age distribution of women in both states is similar to the national average. The two states differ slightly in their race/ethnic compositions. In Maryland a greater share of women were Black or Hispanic. In 2017, 31.2% of women were Non-Hispanic Black and 21.3% were Hispanic. In 2017 in Delaware, these numbers were 22.7% and 16.4% respectively. Delaware women were less likely to be college graduates than were Maryland women (around 35% versus around 45%), and were more likely to live in a household with income under $50,000 (around a third in Delaware versus around a quarter in Maryland). In both states and periods, approximately 40% of women were married, 40% were single, with the remaining 20% cohabiting. In both states and time points, approximately half of women had experienced at least one birth and half were nulliparous. Between five and eight percent of women rated their health as only 'Fair' or 'Poor'. Difficulty accessing birth control in the past 12 months was experienced by about 10 percent of women. Difficulty either accessing birth control or health-care generally in the past 12 months was experienced by between 23 and 29 percent of women. For most characteristics considered, therefore, distributions are reasonably similar between Delaware and Maryland and relatively unchanging over time. This suggests that our not having included more than age and state controls in our regressions is unlikely to have contributed to substantial differences affecting coefficient comparisons between Delaware and Maryland, as might arise were there large differences in distributions of confounders.

[TABLE 3 ABOUT HERE]

Table 3 shows the regression coefficients corresponding to differences in cumulative lifetime and five-year abortion incidence associated with eight different categories of socio-demographic and health or health-care access characteristics. Following from the linear and quadratic age assessments of Figures 2 and 3 respectively, the regression models pooling observations across the two states include a single linear coefficient for age for the cumulative lifetime incidence model and a linear and squared age coefficient for the five-year incidence model. An additional indicator variable for (Maryland) state of residence is also included among the regressors. Because they are from a linear model, the coefficients multiplied by 100 can be interpreted as the percentage-point difference in abortion incidence from that of the reference category for that variable. In particular, the coefficients for 'Maryland' state of residence show magnitudes of abortion incidence that are all between 4 and 8 percentage points higher for Maryland residents than for Delaware residents, although these coefficient estimates vary in their statistical significance. Because we have centered age (at age 25) in these models, the constant term is also informative. It represents the level of abortion incidence, respectively cumulative to age 25 and in the past five years up to age 25, for Delaware women at the reference category of each of the socio-demographic or health characteristics. Thus, on average, 9.7% of Delaware women in 2017 had already had an abortion by age 25, and 5.9% of Delaware women aged 25 in 2021 had an abortion in the past five years.

Differences in incidence are generally larger for the cumulative lifetime incidence than they are for the five-year incidence. This corresponds to the larger overall magnitude of the metric of cumulative lifetime incidence compared to five-year incidence (see again Figure 2 versus Figure 3). As a consequence, and despite sample size being two and a half times larger for the five-year outcome than for the cumulative-lifetime outcome, statistical power to detect differences is generally weaker for the five-year incidence outcome.

The directions of socio-demographic disparities in abortion incidence found in national-level studies using the APS abortion numerators with population denominators (R. K. Jones & Jerman, 2017; R. Jones & Kavanaugh, 2011) are generally replicated here in our Delaware and Maryland estimates. Relative to non-Hispanic White adults, Black adults were 20.3 percentage points more likely to have had an abortion already in their lifetimes (through the survey year 2017) and 12.6 percentage points more likely to have had an abortion in the past five years (to survey year 2021). Individuals living in households with annual incomes less than $50,000 at the time of the interview were 14.0 percentage points more likely to have had an abortion during their lifetimes, and 5.4 percentage points more likely to have had an abortion in the past five years (significant only at the .10 level) compared to those living in households with annual incomes $50,000 or more. Individuals without a four-year college degree were 14.9 percentage points more likely to have had an abortion during their lifetime through the survey year 2017 compared to non-college graduates. Women who were currently cohabiting at the 2017 survey were 27.5 percentage points more likely to have had an abortion during their lifetimes, and 9.5 percentage points more likely during the past five years, than were currently-married women.

Another variable for which the regression coefficients may be compared to previous findings from the APS-based estimates is parity. Our results are again consistent with the directions of those previous findings at the national level: nulliparous women were 16.0 percentage points less likely to have had an abortion during their lifetimes up to 2017 and 10.2 percentage points less likely to have had an abortion in the past five-years up to 2021. Importantly, however, our estimates control for age, which would be expected to be an important confounder. (In results not shown, we ran an equivalent model without age controls and found higher "nulliparous" coefficient magnitudes than these results presented in Table 3 that control for age.)

Also estimated in our regressions are abortion-incidence differences by self-reported heath and by having experienced difficulties either in accessing any health care or in accessing birth control in the last 12 months. Women with fair or poor self-rated health were 26.2 percentage points more likely to have had an abortion during their lifetimes up to 2017 and 11.0 percentage points more likely to have had an abortion in the past five-years up to 2021. Having recent (in the last 12 months) difficulty in accessing birth control, or in accessing health care or birth control, was associated with higher five-year abortion incidence. Women who had difficulty accessing birth control were 9.9 percentage points more likely to have had an abortion in the past five-years to 2021 ($p < .10$), and women were 7.6 percentage points more likely to have had an abortion in the past five-years if they had difficulty accessing birth control or health care more generally ($p < .05$). Recent experience of accessing health care would be expected to be more relevant for abortion incidence in the past five years than for abortion incidence over their lifetimes, and indeed neither of these health-care access measures was statistically-significantly associated with cumulative lifetime abortion incidence.

Discussion

In this study, we used the list experiment method to estimate five-year abortion incidence and cumulative-lifetime abortion incidence using state-representative survey data from Delaware and Maryland. This follows up two earlier studies using population representative data respectively for cumulative-lifetime abortion incidence in Delaware and Maryland (Kissling & Jackson, 2022), and in Ohio (Hood et al., 2022). The present study includes several features that both validate the list-experiment method and extend its utility for analyzing abortion incidence in population-representative U.S. data.

First, we considered abortion incidence over a limited recent period: the past five years. Often researchers' objectives include understanding of recent trends rather than of lifetime experience. To our knowledge, this is the first time that the abortion list experiment has been applied to estimates of abortion over a short period. We found that with a moderate-to-large sample size (around 7,000 women at reproductive ages), estimating five-year incidence with the list experiment was feasible and useful. The declines in statistical power compared with lifetime incidence, resulting from the smaller outcome magnitudes, were nevertheless substantial. These results do not suggest promise, therefore, for use of the list-experiment method for a shorter time interval, for example for the estimation of annual abortion incidence, unless very large sample sizes would be obtained. Because the abortion incidence of Maryland is among the highest nationwide, the present analysis of five-year incidence is likely to be a good test of the limits of short-period abortion incidence estimation using the list-experiment method.

Our use of the five-year incidence outcome metric also facilitated comparison of list experiment results against estimates from external population counts data sources, specifically external population estimates of five-year abortion incidence in Delaware and Maryland calculated from APC data on abortion numbers by state by year, together with data of age distributions and of first versus subsequent abortion data reported in CDC surveillance data. To our knowledge, this is the first time that the abortion list experiment method has been evaluated against benchmark estimates of the true level of abortion incidence. For both Delaware and Maryland, our five-year incidence list-experiment estimates were reasonably close to our external-data benchmark estimates, and the list-experiment confidence intervals included the external-data benchmark estimates in both states. List experiments have in general been evaluated by comparing direct question estimates with those from the list experiment (Ehler et al., 2021; Li & Van den Noortgate, 2022; Hood 2022), with the expectation that estimates would be higher than for the direct question. We were able here to estimate abortion incidence using five years of data from external sources using assumptions to which we believe the estimates are likely to be relatively insensitive over this relatively short period. Notably, we assume the five-year period reported on in the list experiment was lived in the woman's current state of residence. The longer the retrospective period of the incidence reporting, for example when using lifetime incidence as the metric, the more challenging it would be to construct valid benchmark estimates to use in evaluation of list-experiment estimates, because of potential bias arising from inter-state migration.

We were also able to evaluate our list-experiment estimates against several other sources, and each time the list-experiment method results were overall reassuring. Because we used a five-year incidence outcome, we were able to compare our estimates of the age pattern of five-year abortion incidence with both CDC surveillance data (Kortsmit, 2021) and APS data for three recent years, most recently for the 2014 year (Jones and Jerman 2017). As for these external sources, our estimated age pattern from our list-experiment method follows the general pattern of rising then falling risk by age. However, our estimated five-year abortion incidence age peak, at ages 30-34, is somewhat older than in those external sources. The recent CDC compilations, up to 2019 (Kortsmit, 2021), find that abortion incidence peaks at the 20-24 and 25-29 age groups both nationally and in Delaware (no data are available from this source for Maryland). The age pattern of abortion incidence nationally using the APS is even younger, peaking at the 20-24 age group. Even accounting for the five-year look-back of our estimates, and therefore expected older age pattern that for an annual estimation, our estimated age pattern from our list-experiment method is somewhat older than those from both the CDC surveillance data and the APS data.

We were also able to evaluate the Maryland versus Delaware levels of lifetime abortion incidence from our list experiment estimates for two real cohorts (attaining age 40-44 respectively in 2017 and 2021) to the recent annual abortion rates for the two states. The annual abortion rates of Maryland, as calculated from the Guttmacher APC data, have been consistently higher than the overall national rates. Consistent with this, our estimate of lifetime abortion incidence for real cohorts of women in Maryland (34.6% in 2017 and 36.3% in 2021) are higher than found for the most recent estimates of lifetime abortion incidence for synthetic cohorts, including the 24% national estimate using data for the most recent, 2014 Abortion Patient Survey (APS) (Jones & Jerman, 2017) and the 30.1% national estimate using data from the 2008 APS (R. Jones & Kavanaugh, 2011). Our estimates of cumulative lifetime abortion incidence at ages 40-44 in Delaware (27.1% in 2017 and 28.2% in 2021), however, are more similar to these national synthetic-cohort estimates. The closer level for Delaware from our list experiment estimates are not surprising given that abortion rates in Delaware have recently been approaching the national average (see again Figure 1 and (R. Jones et al., 2022)). The abortion rate in

Maryland, meanwhile, has remained much higher than the national average (Maddow-Zimet & Kost, 2021), and has even increased between 2019 and 2020 (R. Jones et al., 2022).

We were also able to compare our list-experiment's cumulative-lifetime and five-year estimates of abortion-incidence differentials by socio-demographic characteristics to APS-based national annual abortion incidence estimates of differentials for 2000, 2008, and 2014 (R. K. Jones & Jerman, 2017; R. Jones & Kavanaugh, 2011). While the magnitude of differences changed across years, in all cases: Black, lower-income, less educated (less than a four year college degree), cohabiting, and women with at least one birth, had higher abortion incidence in both our list-experiment state estimates and in the APS-based national estimates.

The list experiment allows for two important methodological innovations relative to estimates based on APS data. First, the list-experiment allows estimation of abortion incidence by a broader range of characteristics, as no external population denominator data source is required (as it is for APS-based estimation). We were thus able to obtain abortion incidence estimates by women's experiences of health-care access difficulties, which has been proposed as a theoretically important mechanism (Dehlendorf et al., 2013). Similarly, we were able to estimate abortion incidence by self-rated health, with the unsurprising but nevertheless informative finding that women with lower self-rated health had higher abortion incidence, both cumulative and recent.

Second, the list-experiment allows for multivariate estimation. This allowed us to examine associations net of age and state of residence. We controlled for age and state of residence when estimating the incidence by race, income, education, relationship status, and parity, and the health and health-access variables. Age is a common confounder, for example nulliparous women are on average younger than parous women (Mathews & Hamilton, 2016), and therefore have had less time in their lives to have experienced and abortion. We unsurprisingly found that after controlling for age, the 'nulliparous' coefficient was less negative in our regression for lifetime abortion incidence, but not substantially changed for five-year incidence (results not shown).

Our use of multivariate regression estimation is the first that we know of in the U.S. for list-experiment estimation of abortion. Multivariate regression estimation for list-experiment estimation of abortion has previously been successfully applied in a developing-country context (Moseson et al., 2017), but has had mixed success in studies using the list experiment to measure other sensitive topics in the U.S. and elsewhere (Wolter & Laier, 2014). Our use of multivariate regression estimation for five-year incidence in addition to of cumulative-lifetime abortion incidence was also successful for estimating socio-demographic and health differentials. In general, we found that coefficients for five-year incidence were in the same direction as those of lifetime incidence. They were, however, generally lower in magnitude (reflecting lower overall incidence over five years than a lifetime) and more likely to be not statistically significant. The latter was despite sample sizes more than twice as large for the five-year model as for the cumulative lifetime incidence model. A notable exception, however, was experience of recent difficulty accessing health care, for which estimated magnitudes of association with abortion were of higher magnitude and were statistically significant for the five-year incidence measure only. This suggests promise for additional consideration of contextual variables when using a five-year abortion-incidence measure.

In summary, we were able to provide substantial evidence in support of the validity of the list experiment as implemented in two different survey years in the states of Delaware and Maryland, and were additionally able to demonstrate additional utility of the list experiment over that shown in previous applications of the method in U.S. population-representative samples (Hood et al., 2022; Kissling & Jackson, 2022). This utility included the application of the list experiment to shorter time periods (five years), the use of multivariate regression estimation, and the incorporation of a more varied set of predictor variables than those for standard socio-demographic characteristics.

These very promising findings for the abortion list experiment are particularly notable in the context of longstanding problems with survey direct question approaches to understanding abortion that have found abortion incidences markedly lower than levels obtained from external count sources (Lindberg et al., 2020), even in states relatively favorable towards abortion (Maddow-Zimet et al., 2021). There are numerous studies finding large underreporting of abortion in direct question survey self-reports (Jagannathan, 2001; E. F. Jones & Forrest, 1992; R. K. Jones & Kost, 2007; Lindberg et al., 2020; Tierney, 2019). The findings of these studies calling into question the validity of direct questions, moreover, appear to have been taken to heart by researchers. We have been able to find only two published studies (Gius, 2007; Sutton et al., 2019) that have made substantive analytic use of any of the three main nationally-representative surveys' abortion self-report data that use direct questions (the NSFG, Add Health, and NSLY). We argue that the list-experiment method promises to break this unfortunate state of data resources for the study of abortion (Ahrens & Hutcheon, 2020).

Still, a few limitations are of note. In this first application of multivariate regression, we have not fully controlled for other variables to estimate independent effects or associations. For example, it is unclear whether the difference in abortion incidence by relationship status or self-rated health reflects that women with lower socio-economic status are generally more likely to be cohabiting (Cherlin, 2020) and in poorer health (Glymour et al., 2014), or whether cohabiting women or women with health problems are independently more likely to seek abortion services. The relatively low statistical efficiency of the list-experiment method prevented us from performing such analyses in the present study. With a larger survey sample, it may be feasible, and would be of considerable substantive interest, to examine whether such differences in abortion incidence persist after adjusting for socio-economic status as well as age.

Second, for our validation to external data sources of abortion levels, we estimated benchmark abortion incidence from external data sources for women residing in the two states, Delaware and Maryland, whereas some of the women in our sample may have moved from other states with different abortion rates over the reporting period. This is likely to be of less concern over a five-year reporting period, but will be a more challenging issue for assessing the validity of cumulative lifetime abortion incidence against external benchmark estimates.

The issue of the substantially higher variance of list-experiment than direct question methods can be addressed in several ways, including using the double-list variant as we do here (Blair et al., 2020). Pooling observations across states, using a single parameterization of age across both states, was a second method we were able to use. Nevertheless, our estimates of five-year incidence in particular meant that we were not as able to distinguish between sub-groups of interest as we were when estimating differences in cumulative lifetime abortion incidence. This issue could be addressed in future studies with the use of larger samples, by use of statistical techniques that reduce variance with maximum likelihood estimation (Blair & Imai, 2012), and potentially with the use of auxiliary population data (Chou et al., 2020) such as those data available in the APC and state health department compilations provided to the CDC.

Additional work is also needed to evaluate the generalizability of the list experiment approach to other U.S. states. While our study suggested that the list experiment yielded valid inferences in Delaware and Maryland, in line with both external count data sources and with APS-based survey estimates at the national level, other work has concluded less favorably about the list-experiment approach using data for Ohio (Hood et al., 2022). Moreover, two recent meta-analyses of the list-experiment applied to other topics conclude that while list experiments overall may increase estimates of sensitive items, some studies do not find any gains from the list experiment method (Ehler et al., 2021; Li & Van den Noortgate, 2022). Still, the estimates presented here offer the potential for a marked improvement to that of using direct survey questions on abortion, which have consistently failed to yield credible estimates of abortion incidence despite several decades of ongoing efforts to overcome the deficiencies of direct question approaches (E. F. Jones & Forrest, 1992; Lindberg et al., 2020).

References

Ahlquist, J. S. (2018). List experiment design, non-strategic respondent error, and item count technique estimators. *Political Analysis*, *26*(1), 34–53.

Ahrens, K. A., & Hutcheon, J. A. (2020). Time for better access to high-quality abortion data in the United States. *American Journal of Epidemiology*, *189*(7), 640–647. https://doi.org/10.1093/aje/kwaa048

Bell, S. O., & Bishai, D. (2019). Can a list experiment improve validity of abortion measurement? *Studies in Family Planning*, *50*(1), 43–61. https://doi.org/10.1111/sifp.12082

Blair, G., Chou, W., & Imai, K. (2019). List experiments with measurement error. *Political Analysis*, *27*(4), 455–480.

Blair, G., Coppock, A., & Moor, M. (2020). When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *American Political Science Review*, *114*(4), 1297–1315.

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, *20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Boudreaux, M., & Rendall, M., S. (2020). *Statewide Survey of Women of Reproductive Age in Delaware and Maryland Baseline Survey [computer file] [Data set]*. https://popcenter.umd.edu/delcaneval/survey

Boudreaux, M., Eeckhaut, M., & Rendall, M., S. (2022). *Statewide Survey of Women of Reproductive Age in Delaware and Maryland Endline Survey [computer file] [Data set]*. https://popcenter.umd.edu/delcaneval/endline

Center for Disease Control. (2022). *2017-2019 NSFG user's guide appendix 2: topic-specific notes for 2017-2019. https://www.cdc.gov/nchs/nsfg/nsfg_2017_2019_puf.htm*. Center for Disease Control. https://www.cdc.gov/nchs/nsfg/nsfg_2017_2019_puf.htm

Cherlin, A. J. (2020). Degrees of change: An assessment of the deinstitutionalization of marriage thesis. *Journal of Marriage and Family*, *82*(1), 62–80.

Chou, W., Imai, K., & Rosenfeld, B. (2020). Sensitive survey questions with auxiliary information. *Sociological Methods & Research*, *49*(2), 418–454.

Cohen, I. G., Murray, M., & Gostin, L. O. (2022). The end of Roe v Wade and new legal frontiers on the constitutional right to abortion. *Journal of the American Medical Association (JAMA)*, (328)(4), 325-326 doi:10.1001/jama.2022.12397

Cowan, S. K., Wu, L. L., Makela, S., & England, P. (2016). Alternative estimates of lifetime prevalence of abortion from indirect survey questioning methods. *Perspectives on Sexual and Reproductive Health*, *48*(4), 229–234. https://doi.org/10.1363/48e11216

Dehlendorf, C., Harris, L. H., & Weitz, T. A. (2013). Disparities in abortion rates: A public health approach. *American Journal of Public Health*, *103*(10), 1772–1779.

Desai, S., Lindberg, L. D., Maddow-Zimet, I., & Kost, K. (2021). The impact of abortion underreporting on pregnancy data and related research. *Maternal and Child Health Journal*, *25*(8), 1187–1192.

Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., Ezzati, T. M., Biemer, P. P., Groves, R. M., Lyberg, L. E., & Mathiowetz, N. A. (1991). Measurement errors in surveys. *The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application*, 185–210.

Ehler, I., Wolter, F., & Junkermann, J. (2021). Sensitive questions in surveys: A comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. *Public Opinion Quarterly*, *85*(1), 6–27.

Forrest, J. D. (1987). Unintended pregnancy among American women. *Family Planning Perspectives*, *19*(2), 76–77.

Gius, M. P. (2007). The impact of provider availability and legal restrictions on the demand for abortions by young women. *The Social Science Journal*, *44*(3), 495–506.

Glymour, M. M., Avendano, M., & Kawachi, I. (2014). Socioeconomic status and health. In L. F. Berkman, I. Kawachi, & M. M. Glymour (Eds.), *Social Epidemiology* (p. 0). Oxford University Press. https://doi.org/10.1093/med/9780195377903.003.0002

Hood, R. B., Moseson, H., Smith, M., Chakraborty, P., Norris, A. H., & Gallo, M. F. (2022). Comparison of abortion incidence estimates derived from direct survey questions versus the list experiment among women in Ohio. *PLOS ONE*, *17*(6), e0269476. https://doi.org/10.1371/journal.pone.0269476

Imai, K., Park, B., & Greene, K. F. (2015). Using the predicted responses from list experiments as explanatory variables in regression models. *Political Analysis*, *23*(2), 180–196.

Jagannathan, R. (2011). Relying on surveys to understand abortion behavior: Some cautionary evidence. *American Journal of Public Health*, *91*(11), 1825–1831.

Jones, E. F., & Forrest, J. D. (1992). Underreporting of abortion in surveys of US women: 1976 to 1988. *Demography*, *29*(1), 113–126.

Jones, R., Jerman, J., & Ingerick, M. (2018). Which Abortion Patients Have Had a Prior Abortion? Findings from the 2014 U.S. Abortion Patient Survey. *Journal of Women's Health*, *27*(1), 58–63. https://doi.org/10.1089/jwh.2017.6410

Jones, R. K., & Jerman, J. (2017). Population group abortion rates and lifetime incidence of abortion: United States, 2008–2014. *American Journal of Public Health*, *107*(12), 1904–1909.

Jones, R. K., & Kost, K. (2007). Underreporting of induced and spontaneous abortion in the United States: An analysis of the 2002 National Survey of Family Growth. *Studies in Family Planning*, *38*(3), 187–197.

Jones, R., & Kavanaugh, M. (2011). Changes in abortion rates between 2000 and 2008 and lifetime incidence of abortion. *Obstetrics & Gynecology*, *117*(6), 1358–1366.

Jones, R., Philbin, J., Kirstein, M., Nash, E., & Lufkin, K. (2022, June 13). *Long-Term Decline in US Abortions Reverses, Showing Rising Need for Abortion as Supreme Court Is Poised to Overturn Roe v. Wade*. Guttmacher Institute. https://www.guttmacher.org/article/2022/06/long-term-decline-us-abortions-reverses-showing-rising-need-abortion-supreme-court

Kissling, A., & Jackson, H. M. (2022). estimating prevalence of abortion using list experiments: findings from a Survey of Women in Delaware and Maryland. *Women's Health Issues*, *32*(1), 33–40. https://doi.org/10.1016/j.whi.2021.08.003

Kortsmit, K. (2021). Abortion surveillance—United States, 2019. *MMWR. Surveillance Summaries*, *70*. https://doi.org/10.15585/mmwr.ss7009a1

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial attitudes and the "New South". *The Journal of Politics*, *59*(2), 323–349.

Lara, D., Strickler, J., Olavarrieta, C. D., & Ellertson, C. (2004). Measuring induced abortion in Mexico: a comparison of four methodologies. *Sociological Methods & Research*, *32*(4), 529–558. https://doi.org/10.1177/0049124103262685

Lépine, A., Treibich, C., & D'Exelle, B. (2020). Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours. *Social Science & Medicine*, *266*, 113326. https://doi.org/10.1016/j.socscimed.2020.113326

Li, J., & Van den Noortgate, W. (2022). A Meta-analysis of the relative effectiveness of the item count technique compared to direct questioning. *Sociological Methods & Research*, *51*(2), 760–799.

Lindberg, L., Kost, K., Maddow-Zimet, I., Desai, S., & Zolna, M. (2020). Abortion reporting in the United States: An assessment of three national fertility surveys. *Demography*, *57*(3), 899–925. https://doi.org/10.1007/s13524-020-00886-4

Maddow-Zimet, I., & Kost, K. (2021). *Pregnancies, births and abortions in the United States, 1973–2017: National and state trends by age*.

Maddow-Zimet, I., Lindberg, L. D., & Castle, K. (2021). State-Level Variation in Abortion Stigma and Women and Men's Abortion Underreporting in the USA. *Population Research and Policy Review*, *40*(6), 1149–1161. https://doi.org/10.1007/s11113-021-09657-4

Mathews, T. J., & Hamilton, B. E. (2016). Mean age of mothers is on the rise: United States, 2000–2014. NCHS data brief, no 232. *Hyattsville, MD: National Center for Health Statistics*.

Moseson, H., Filippa, S., Baum, S. E., Gerdts, C., & Grossman, D. (2019). Reducing underreporting of stigmatized pregnancy outcomes: Results from a mixed-methods study of self-managed abortion in Texas using the list-experiment method. *BMC Women's Health*, *19*(1), 113. https://doi.org/10.1186/s12905-019-0812-4

Moseson, H., Gerdts, C., Dehlendorf, C., Hiatt, R. A., & Vittinghoff, E. (2017). Multivariable regression analysis of list experiment data on abortion: Results from a large, randomly-selected population based study in Liberia. *Population Health Metrics*, *15*(1), 40. https://doi.org/10.1186/s12963-017-0157-x

Moseson, H., Massaquoi, M., Dehlendorf, C., Bawo, L., Dahn, B., Zolia, Y., Vittinghoff, E., Hiatt, R. A., & Gerdts, C. (2015). Reducing under-reporting of stigmatized health events using the List Experiment: Results from a randomized, population-based study of abortion in Liberia. *International Journal of Epidemiology*, *44*(6), 1951–1958. https://doi.org/10.1093/ije/dyv174

National Opinion Research Center. (2019). *Methodology Report: Delaware Contraceptive Access Now (DelCAN): Baseline Survey* (pp. 1–34). NORC at the University of Chicago. https://popcenter.umd.edu/delcaneval/files/SoW_Baseline_Survey_Methodology_202005

National Opinion Research Center. (2021). *Methodology Report: Delaware Contraceptive Access Now (DelCAN): Endline Survey* (pp. 1–35). NORC at the University of Chicago. https://popcenter.umd.edu/delcaneval/files/endline_methodology_rpt_2021.pdf

Solazzo, A. L. (2019). Different and not equal: The uneven association of race, poverty, and abortion laws on abortion timing. *Social Problems*, *66*(4), 519–547.

Stillman, M., Leong, E., Utomo, B., Dadun, D., Aryanty, R. I., Sedgh, G., & Giorgio, M. M. (2020). An application of the confidante method to estimate induced abortion incidence in Java, Indonesia. *International Perspectives on Sexual and Reproductive Health*, *46*, 199–210. https://doi.org/10.1363/46e0120

Sully, E., Giorgio, M., & Anjur-Dietrich, S. (2020). Estimating abortion incidence using the network scale-up method. *Demographic Research*, *43*, 1651–1684.

Sutton, A., Lichter, D. T., & Sassler, S. (2019). Rural–urban disparities in pregnancy intentions, births, and abortions among US adolescent and young women, 1995–2017. *American Journal of Public Health*, *109*(12), 1762–1769.

Tierney, K. I. (2019). Abortion underreporting in Add Health: Findings and implications. *Population Research and Policy Review*, *38*(3), 417–428.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, *133*(5), 859.

Tsai, C. (2019). Statistical analysis of the item-count technique using Stata. *The Stata Journal*, *19*(2), 390–434. https://doi.org/10.1177/1536867X19854018

Udry, J. R., Gaughan, M., Schwingl, P. J., & Van Den Berg, B. J. (1996). A medical record linkage analysis of abortion underreporting. *Family Planning Perspectives*, 228–231.

Wolter, F., & Laier, B. (2014). The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. *Survey Research Methods*, *8*(3), 153–168.

Yeatman, S., & Trinitapoli, J. (2011). Best-friend reports: A tool for measuring the prevalence of sensitive behaviors. *American Journal of Public Health*, *101*(9), 1666–1667.

Table 1. Double List Experiment Items Administered in Baseline and Endline Delaware and Maryland Survey of Women

On the following list of health experiences, how many of these have you personally experienced? You don't need to say which ones, just how many.	
Version A- Control	• Ever used or taken medication for which a prescription is needed • Ever had a pap smear • Diagnosed with breast cancer in the past 10 years
Version A- Treatment	• [Baseline Only] Ever had an abortion (ended a pregnancy on purpose) • [Endline Only] Had an abortion (ended a pregnancy on purpose) in the past 5 years • Ever used or taken medication for which a prescription is needed • Ever had a pap smear • Diagnosed with breast cancer in the past 10 years
Version B- Control	• Ever used a birth control method (such as: pills, an IUD or implant, condoms or the shot) • Had a tubal or ectopic pregnancy in the past year • Ever had your blood pressure measured
Version B- Treatment	• [Baseline Only] Ever had an abortion (ended a pregnancy on purpose) • [Endline Only] Had an abortion (ended a pregnancy on purpose) in the past 5 years • Ever used a birth control method (such as: pills, an IUD or implant, condoms or the shot) • Had a tubal or ectopic pregnancy in the past year • Ever had your blood pressure measured

Table 2. Characteristics of Women Aged 18-44 Living in Delaware and Maryland in 2017 and 2021, proportions

	2017		2021	
	Maryland N=1,349	Delaware N=1,398	Maryland N=3,097	Delaware N=4,123
Age				
18-24	0.231	0.246	0.226	0.228
25-29	0.173	0.194	0.187	0.202
30-34	0.210	0.193	0.203	0.206
35-39	0.202	0.167	0.200	0.172
40-44	0.184	0.201	0.183	0.192
Race				
Non-Hispanic White	0.475	0.608	0.455	0.563
Non-Hispanic Black	0.312	0.227	0.326	0.249
Hispanic or Other Race/Ethnicity	0.213	0.164	0.218	0.188
Household Income				
Below $50,000	0.268	0.366	0.239	0.309
At or Above $50,000	0.732	0.634	0.761	0.691
Education				
Less than High School	0.029	0.051	0.022	0.029
High School	0.119	0.167	0.143	0.169
Some College	0.403	0.441	0.388	0.436
College Graduate	0.449	0.341	0.448	0.366
Partnership Status				
Married	0.421	0.392	0.420	0.415
Cohabiting	0.182	0.227	0.178	0.195
Single	0.397	0.381	0.402	0.390
Parity				
Parous (1+ Births)	0.514	0.526	0.480	0.497
Nulliparous (0 Births)	0.486	0.474	0.520	0.503
Self Rated Health				
Excellent, Very Good or Good	0.952	0.924	0.924	0.923
Fair or Poor	0.048	0.076	0.076	0.077
Health Care Access in Past 12 Months				
No Difficulty Accessing Birth Control	0.899	0.871	0.901	0.896
Difficulty Accessing Birth Control	0.101	0.129	0.099	0.104
No Difficulty Accessing Birth Control or Health Care	0.767	0.741	0.711	0.706
Difficulty Accessing Birth Control or Health Care	0.233	0.259	0.289	0.294

Notes: Age, race, income, education and marital status have some values imputed by the survey provider.
Source: 2017 Baseline and 2021 Endline Survey of Women in Delaware and Maryland

Table 3. Linear Regression Coefficients for Characteristics Associated with Cumulative Lifetime and Five-Year Abortion Incidence, Women Aged 18-44 Living in Delaware and Maryland 2017 and 2021

	Overall		Race		Income		Education	
	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year
Ref. Delaware								
Maryland	0.071+	0.043	0.050	0.035	0.083+	0.047+	0.084*	0.046
	(0.043)	(0.028)	(0.043)	(0.027)	(0.043)	(0.028)	(0.043)	(0.028)
Age	0.011***	0.007	0.010**	0.006	0.012***	0.007	0.013***	0.008+
	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)
Age Squared		-0.000		-0.000		-0.000		-0.001+
		(0.000)		(0.000)		(0.000)		(0.000)
Ref. White Non-Hispanic								
Black Non-Hispanic			0.203***	0.126***				
			(0.058)	(0.035)				
Hispanic or Other Race/Ethnicity			0.029	-0.028				
			(0.059)	(0.037)				
Ref. Household Income at Least $50,000								
Household Income Less than $50,000					0.140**	0.054+		
					(0.053)	(0.032)		
Ref. Four Year College Degree or Higher								
Less than Four Year College Degree							0.149***	0.031
							(0.041)	(0.026)
Constant	0.097**	0.059*	0.053	0.036	0.042	0.040	-0.015	0.032
	(0.037)	(0.023)	(0.037)	(0.023)	(0.042)	(0.026)	(0.040)	(0.026)
Observations	2705	6941	2705	6941	2705	6941	2705	6941

Table 3 (Continued). Linear Regression Coefficients for Characteristics Associated with Cumulative Lifetime and Five-Year Abortion Incidence, Women Aged 18-44 Living in Delaware and Maryland 2017 and 2021

	Partnership Status		Parity		Self-Rated Health		Birth Control Difficulty		Birth Control or Access Difficulty	
	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year	Cumulative Lifetime	Five-Year
<u>Ref. Delaware</u>										
Maryland	0.083+	0.043	0.069	0.050+	0.078+	0.043	0.068	0.044	0.067	0.045
	(0.042)	(0.027)	(0.043)	(0.028)	(0.043)	(0.027)	(0.042)	(0.028)	(0.043)	(0.028)
Age	0.013***	0.008	0.006	0.004	0.010**	0.007	0.011***	0.007	0.011***	0.006
	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)	(0.003)	(0.005)
Age Squared		-0.000		-0.000		-0.000		-0.000		-0.000
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
<u>Ref. Married</u>										
Cohabiting	0.275***	0.095*								
	(0.058)	(0.038)								
Single	0.070	0.045								
	(0.050)	(0.033)								
<u>Ref. Parous (1+ Births)</u>										
Nulliparous (0 Births)			-0.160***	-0.102**						
			(0.045)	(0.033)						
<u>Ref. Excellent, Very Good, or Good Health</u>										
Fair or Poor Health					0.262*	0.110*				
					(0.105)	(0.054)				
<u>Ref. No Difficulty Accessing Birth Control</u>										
Difficulty Accessing Birth Control							0.079	0.099+		
							(0.069)	(0.051)		
<u>Ref. No Birth Control or Health Care Difficulty</u>										
Difficulty Accessing Birth Control or Health Care									0.049	0.076*
									(0.051)	(0.030)
Constant	-0.007	0.013	0.202***	0.128**	0.080*	0.051*	0.088*	0.046*	0.089*	0.036
	(0.047)	(0.031)	(0.049)	(0.036)	(0.038)	(0.023)	(0.039)	(0.023)	(0.039)	(0.024)
Observations	2705	6941	2496	6362	2693	6937	2690	6923	2682	6911

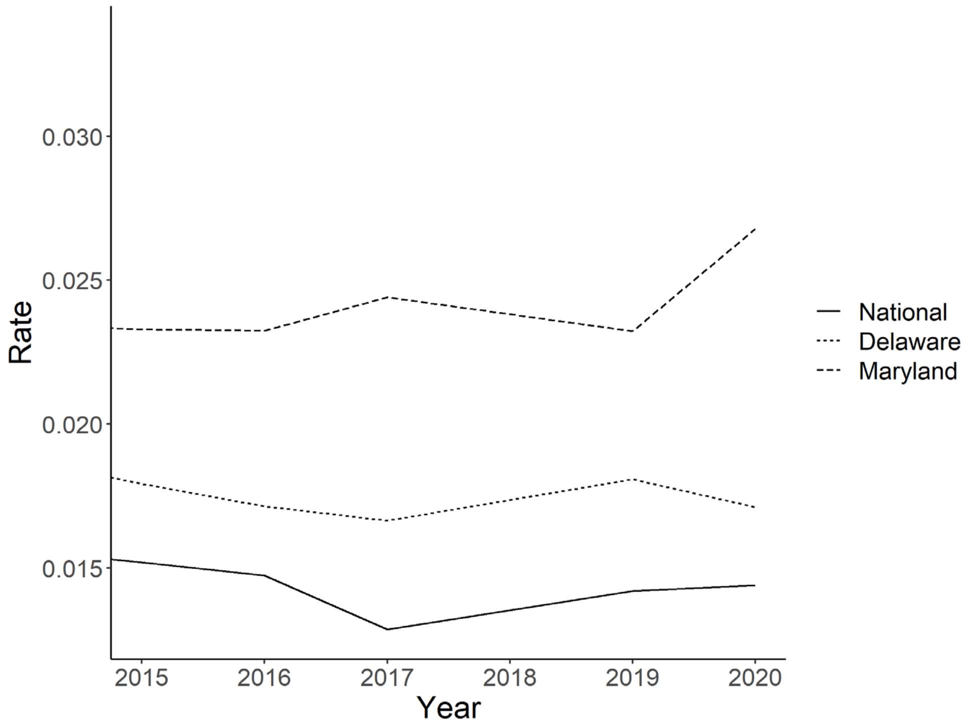
Notes: ‘Age’ and ‘Age Squared’ are ‘Age – 25’ and ‘(Age – 25) Squared’.

+ p<.10 * p<.05 **p<.01 ***p<.001
Standard errors in parentheses.

Coefficients from regression estimate pooling across survey lists using the kict package with linear estimator in Stata (Tsai, 2019). Coefficients multiplied by 100 represent percentage point change in abortion incidence.

Source: 2017 Baseline and 2021 Endline Survey of Women Data in Delaware and Maryland. Baseline survey captures cumulative lifetime abortion incidence; Endline survey captures five-year abortion incidence.

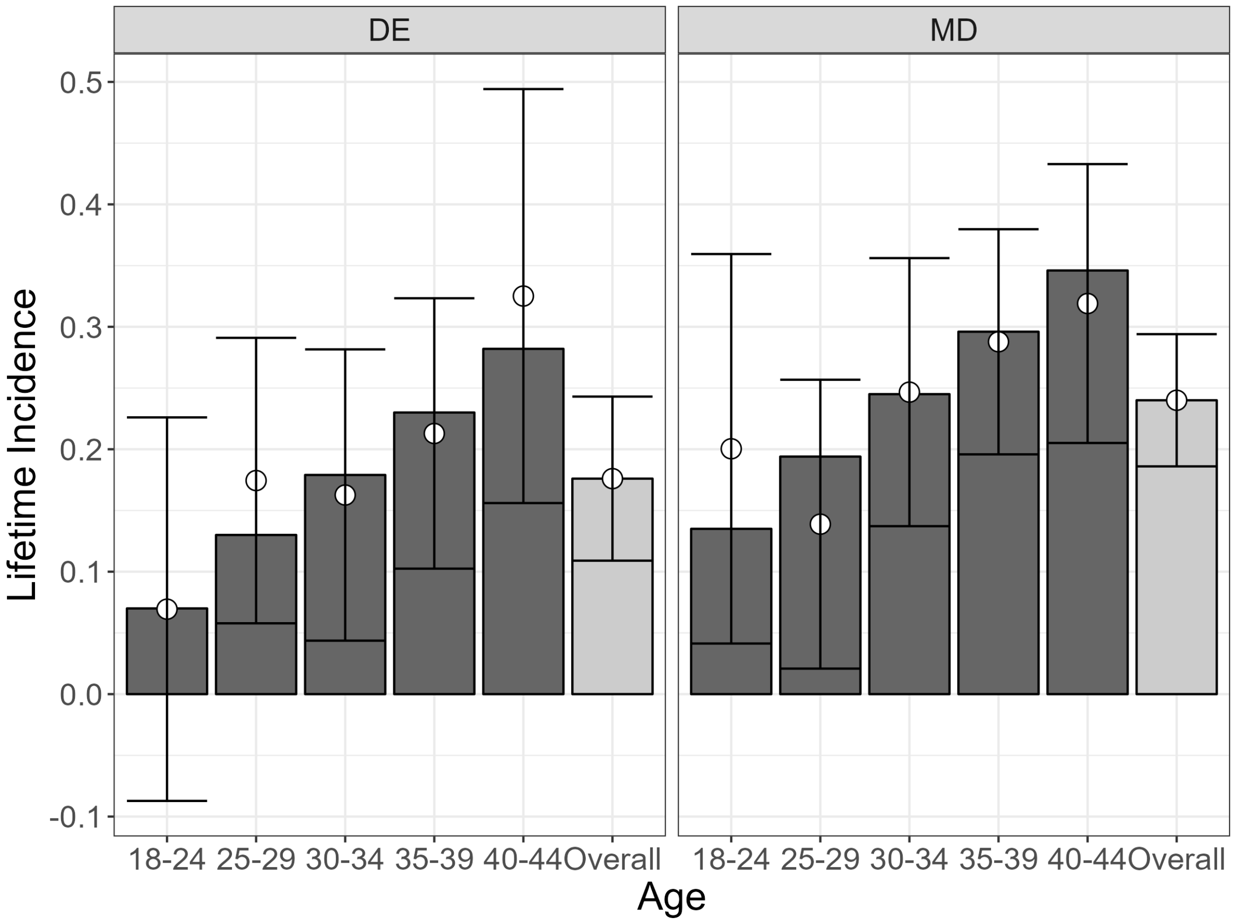
Figure 1 Annual Abortion Rate 2015-2020 among Women 18-44



Notes: Abortion rates directly measured for 2016, 2017, 2019, and 2020.

Source: 2016-2017 Abortion Provider Census Data as published in *Pregnancies, births and abortions in the United States, 1973–2017* (Maddow Zimet & Kost, 2021). Preliminary 2019-2020 Abortion Provider Census Data (R. Jones et al, 2022).

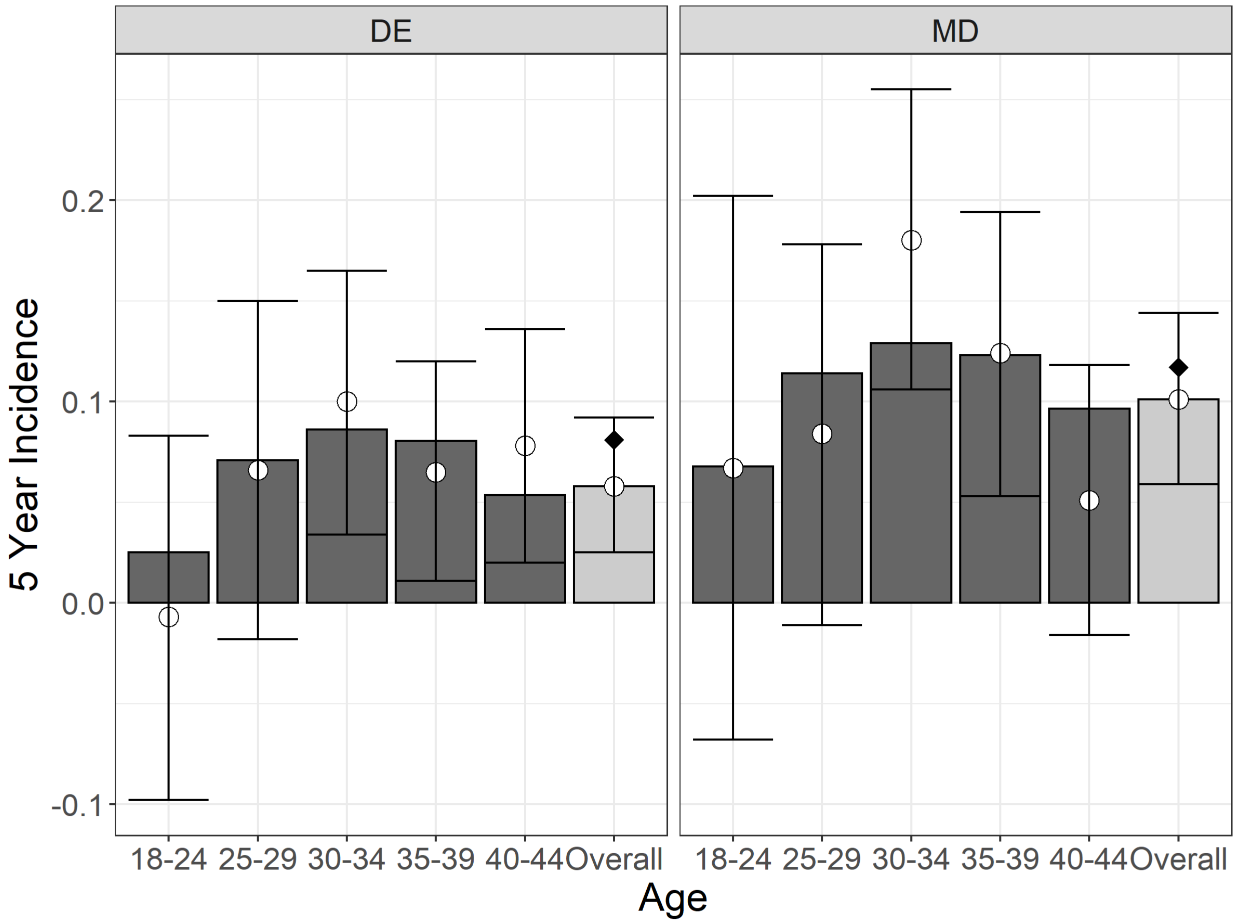
Figure 2: Cumulative Lifetime Abortion Incidence in 2017 by Age by State, Delaware (DE) and Maryland (MD)



Notes: Dark gray bars are predicted values from a regression that pools observations across states and parameterizes age. White points with surrounding confidence intervals are difference-in-means estimates separately by age group and state with the associated 95% confidence intervals. Light gray bars are overall (age 18-44) estimates pooling observations across states.

Source: 2017 Baseline Survey of Women in Delaware and Maryland

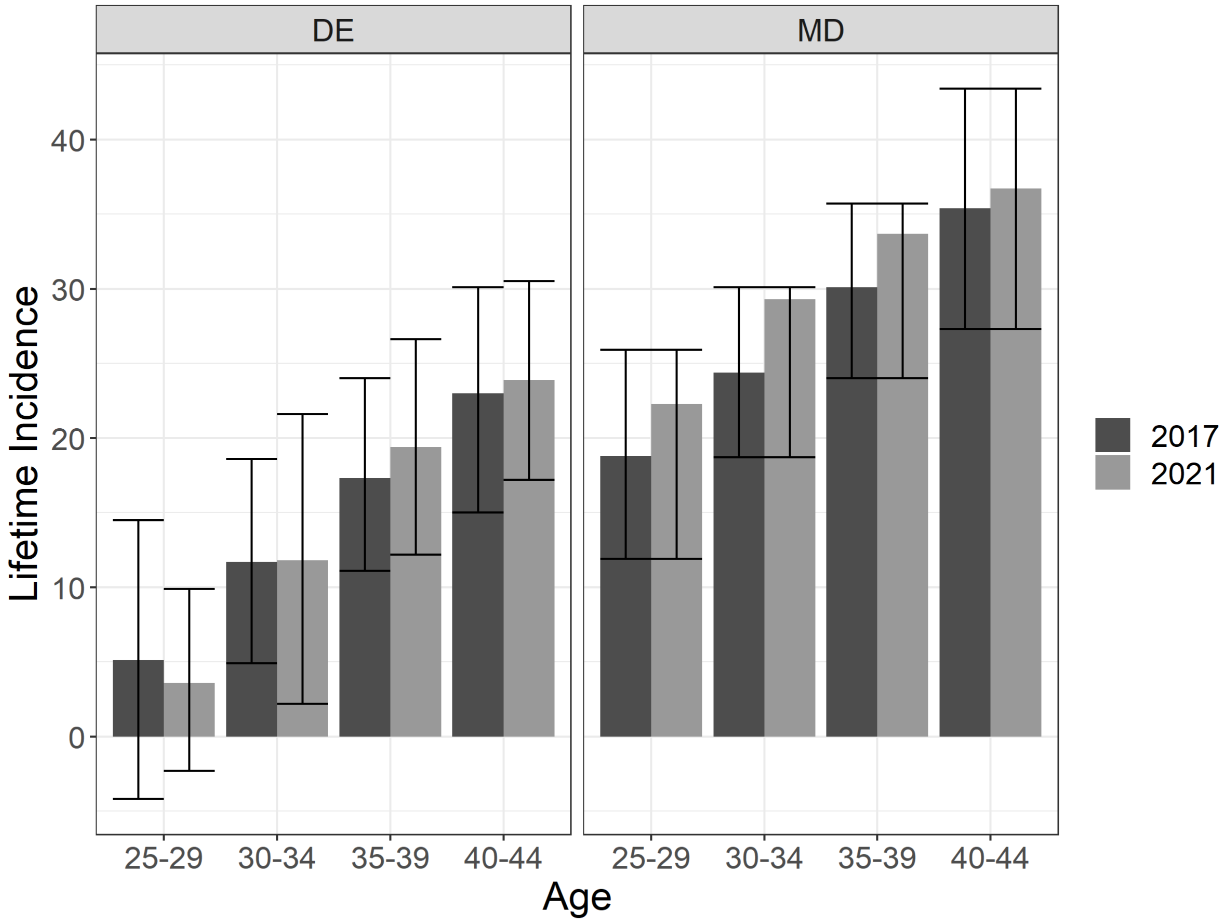
Figure 3: Five-Year Abortion Incidence to 2021 by Age by State, Delaware (DE) and Maryland (MD)



Notes: Dark gray bars are predicted values from a regression that pools across states and parameterizes age. White points with surrounding confidence intervals are difference-in-means estimates separately by age group and state with the associated 95% confidence intervals. Light gray bars are overall (age 18-44) estimates pooling observations across states. Black diamonds are our estimates from external data sources.

Source: 2021 Endline Survey of Women in Delaware and Maryland

Figure 4: Cumulative Lifetime Abortion Incidence in 2017 and 2021 in Delaware (DE) and Maryland (MD)



Notes: Dark gray barplots are 2017 cumulative lifetime incidence estimates predicted from a regression controlling for age and state of residence. Light gray barplots are 2021 lifetime incidence estimates calculated by combining the 2017 cumulative lifetime incidence estimates with five-year incidence estimates predicted from a regression controlling for age, age squared, and state of residence. Error bars are confidence intervals around predicted values taken from 1000 bootstrapped samples.

Source: 2017 Baseline and 2021 Endline Survey of Women in Delaware and Maryland

Methodological Appendix

Appendix 1 Tests of List Experiment Assumptions

In this methodological appendix, we describe the procedures used to confirm list experiment assumptions in the Endline Survey of Women. These same checks are performed on the Baseline Survey with results described in Kissling & Jackson, 2022. First, we confirm that respondents were in fact randomized across lists by examining the state of residence, age, and racial composition of respondents receiving Treatment List A versus Treatment List B and conducting a series of t-tests to test for statistically significant differences across groups. Significant differences across lists may suggest a possible failure to randomize. Next, we use the Blair and Imai design effect test to determine if list experiment response assumptions are met for Lists A and B (Blair & Imai, 2012). The Blair and Imai design effect test examines the cumulative affirmative responses across treatment and control lists and uses a likelihood ratio test to determine if the observed response pattern indicates a failure to reject no design effect. Formally, this test

examines:

$$\Pr(Y_i \ \leq y \mid T_i \ = 0) \geq \Pr(Y_i \ \leq y \mid T_i \ = 1) \ for\ all\ y\ 0, \ldots 3$$

$$\Pr(Y_i \ \leq y \mid T_i \ = 1) \geq \Pr(Y_i \ \leq y - 1 \mid T_i \ = 0) \ for\ all\ y\ 1, \ldots 4$$

such that Y_i indicates the number of items that apply to the respondent, and T_i indicates membership in the treatment group, and y indicates available response options (0-3 for the control list; 0-4 for the treatment list). For list experiment assumptions to be met, the proportion of control list cases agreeing to no more than 0, 1, 2, or 3 items should be greater than the proportion of cases receiving the treatment list, and the proportion of cases in the treatment list who agree to 1, 2, 3, or 4 items should be greater than the proportion of control cases who agree to y-1 items (Blair & Imai, 2012).

P-values from this test have a Bonferroni correction applied (following Blair and Imai) because directly testing the null hypothesis is problematic (Wolak, 1991). As a supplemental check of consistency across lists, we follow (Lépine et al., 2020) and separately estimate the incidence of abortion for each list and use a Wald test to check for differences in the treatment coefficients (capturing incidence). In this test, we separately estimate:

$$Y_i = \beta_A T_i + \varepsilon$$

$$Y_i = \beta_B T_i + \varepsilon$$

Such that Y_i is the number of items that apply to the respondent, T_i is treatment group membership, and β_A is the coefficient (capturing incidence) for list A and β_B is the coefficient for list B. Failure to reject the null hypothesis that $\beta_A = \beta_B$ in the Wald test suggests that incidence is not statistically different across lists and thus lists are internally consistent.

Results

Table S1 shows the weighted characteristics of SoW respondents across treatment groups. T-tests of state of residence, age, race/ethnicity, and number of living children born reveal no statistically significant differences across groups. These findings suggest randomization of the questionnaire was performed as intended. Next, we run the Blair-Imai design effect test in order to examine whether there is evidence that respondents are altering their responses to the control questions when the treatment question is presented. Table S2 shows that for both lists we fail to reject the no-design effect assumption ($p>.05$ for List A and List B); respondents do not appear to be altering their responses on the control questions when they also receive the treatment question on abortion. Finally, we estimated abortion incidence separately by list and found that it did not vary by list ($F=.22$; $p>05$).

Table S1. Characteristics by Treatment Group, Women Aged 18-44 Living in Delaware and Maryland in 2021

	Received Treatment A N=3,711		Received Treatment B N=3,509	
	Proportion[a]	Counts	Proportion[a]	Counts
Delaware	0.556	2079	0.590	2044
Maryland	0.444	1632	0.410	1465
Black Non-Hispanic	0.288	781	0.276	741
Asian Non-Hispanic	0.061	181	0.065	164
Hispanic	0.099	316	0.103	305
Multiple/Other	0.036	121	0.038	123
White Non-Hispanic	0.516	2312	0.518	2176
18-24	0.227	429	0.228	402
25-29	0.194	553	0.197	549
30-34	0.215	843	0.194	732
35-39	0.181	899	0.187	888
40-44	0.183	987	0.194	938
Less than High School	0.025	59	0.026	67
High School	0.145	360	0.172	394
Some College	0.429	1056	0.401	965
College Graduate	0.401	2236	0.401	2083
Household Income Below 25,000	0.136	447	0.134	422
Household Income Between 25,000-75,000	0.266	1291	0.263	1178
Household Income At or Above 75,000	0.597	1973	0.603	1909
Married [b]	0.417	1915	0.418	1852
Cohabiting [b]	0.193	605	0.182	545
Single [b]	0.390	1191	0.400	1112
0 Live Births [b]	0.517	1456	0.503	1347
1 Live Birth [b]	0.173	656	0.184	628
2 Live Births [b]	0.195	817	0.188	749
3 Live Births [b]	0.076	316	0.084	300
4 Live Births [b]	0.038	153	0.042	151

Notes:

a. Proportions calculated using normalized survey weights
b. Counts will not match total sample size due to item-missingness

Source: 2021 Endline Survey of Women in Delaware and Maryland

Table S2. Blair-Imai Design Effect Tests by Treatment List Assignment in 2021

List A

Group	Response to List Experiment				
	0	1	2	3	4
Proportion of Treatment Group Responding	0.052	0.135	0.707	0.092	0.014
Proportion of Treatment Group who Respond at Least	1	0.948	0.812	0.106	0.014
Proportion of Control Group Responding	0.051	0.132	0.784	0.033	0
Proportion of Control Group who Respond at Least	1	0.949	0.817	0.033	0
Treatment – Control Group Cumulative Response	0	-0.001	-0.005	0.072	0.014

Bonferroni Corrected P-Value 0.913

List B

Group	Response to List Experiment				
	0	1	2	3	4
Proportion of Treatment Group Responding	0.039	0.166	0.671	0.106	0.017
Proportion of Treatment Group who Respond at Least	1	0.961	0.795	0.123	0.017
Proportion of Control Group Responding	0.036	0.157	0.765	0.042	0
Proportion of Control Group who Respond at Least	1	0.964	0.807	0.042	0
Treatment – Control Group Cumulative Response	0	-0.003	-0.013	0.081	0.017

Bonferroni Corrected P-Value 0.334

Notes: Following standard practice, design effect tests are unweighted. Results reflect respondent responses and are not generalizable to the population of women 18-44 living in Delaware and Maryland.
Source: 2021 Endline Survey of Women in Delaware and Maryland

Appendix 2 Calculation of 2016-2020 Abortion Incidence for Delaware and for Maryland Using External Population Data

In this supplemental appendix, we detail the methodology used to calculate the five-year abortion incidence in Delaware and Maryland among women aged 18-44 using external counts data. These two calculated numbers (8.1% for Delaware and 11.7% for Maryland) are shown in Figure 3 as black diamonds within the 'Overall age 18-44' bar respectively of the list experiment estimate for Delaware and for Maryland. In both cases, the calculation using these external counts data falls within the list experiment confidence interval, of (2.5-9.2%) for Delaware and of (5.9-14.4%) for Maryland.

We use five-year abortion incidence for women aged 18-44 over calendar years 2016-2020 as the benchmark for the list experiment five-year abortion incidence to Endline Survey date of February 2021 through November 2021. Abortion rates in the two states have been stable enough across the period 2016-2020 to indicate that this is not likely to be an important difference in period.

The most difficult challenge with the external counts data is to exclude abortions that are not the first that the woman has had in the 2016-2020 period, and thereby obtain an estimate of the number of abortions to distinct women aged 18-44 in the 2016-2020 period. This task is made easier by starting by conceptually identifying when an abortion is a first abortion of the 2016-2020 period. First, all abortions of order 1 (a first abortion ever in the woman's lifetime) are by definition her first abortion in the 2016-2020 period. As we show below, approximately 60% of abortions each year are first (ever) abortions. This leaves us with the other 40% of abortions in 2016-2020 that may not have been the woman's first abortion over that five-year period. We next assume that a woman has at most one abortion in a given calendar year. This implies that all abortions in the calendar year 2016 are first abortions in the 2016-2020 period. This leaves us with the task of estimating how many higher-order abortions in 2017-2020 (in each of the two states, DE and MD) are the first abortion the woman has had in the 2016-2020 period. As the time since 2016 increases, the fraction of total abortions that are first abortions over the follow-up period decreases simply because each woman has had increasingly more years to have had a prior abortion within the 2016-2020 period. Our estimated fractions of all abortions that are first abortions that a woman has had during the 2016-2020 period are shown by year and state in Table S3.

Table S3: Percent of Abortions that are First Abortions of the 2016-2020 period, by Year and State

	2016	2017	2018	2019	2020	Overall
Delaware	100.0	96.9	94.3	91.3	88.8	94.2
Maryland	100.0	96.8	94.0	90.9	87.9	94.0

By assumption of no more than one abortion per year, all abortions in calendar year 2016 are first abortions of the 2016-2020 period. In each subsequent year, the fraction of all abortions that are first abortions of the 2016-2020 period naturally declines, from about 97% in 2017 to 88-89% in 2020. Overall, we estimate 94.2% of all abortions in Delaware and 94.0% of all abortions in Maryland in calendar years 2016-2020 are the first abortion a woman has had over the period. Alternately stated, the 94.2 and 94.0 percentages represent the percentage of abortions that occurred to distinct women over the five-year period covered by 2016-2020. These fractions in Table S3 are derived from our use of age and parity distributions of abortions to adjust downwards the number of total abortions in Delaware and in Maryland each calendar year from the Guttmacher Abortion Provider Census (APC) data,(*Guttmacher Data Center*, n.d.;

Maddow-Zimet & Kost, 2021) as we detail below. We sum the distinct abortions from 2016 to 2020 in order to obtain our numerator ('Abortions to Distinct Women 18-44'). The population denominator is the mean population size of women aged 18-44 residing in these two states (from U.S. Census Bureau American Community Survey data(Ruggles, Steven et al., 2022)):

$$2016-2020\ Abortion\ Incidence = \frac{Abortions\ to\ Distinct\ Women\ 18-44}{Mean\ 2016-2020\ Population\ of\ Women\ 18-44}$$

The calculated values of the numerator, denominator, and 2016-2020 abortion incidence percentages for Delaware and Maryland are as follows:

1. Numerators:

$$\sum abortions\ to\ distinct\ women\ per\ year$$

Delaware: 13,214.4

Maryland: 126,668.4

2. Denominators: $\frac{\sum, N_{2016}, N_{2017}, N_{2018}, N_{2019}, N_{2020},}{5}$ where N is the population of women aged 18-44

Delaware: 162,695

Maryland: 1,079,876

3. Calculation of Abortion Incidence in 2016-2020 for Women 18-44

Delaware: 13,214.4/162,695=8.1%

Maryland: 126,668.4/1,079,876=11.7%

Estimating how many higher-order abortions in 2017-2020 are the first abortion the woman has had in the 2016-2020 period

The data used in assigning distributions of abortions by their order within the 2016-2020 period are from distributions of women by age at abortion, combined with distributions of abortions by their lifetime order for the woman. We do not have available to us distributions of abortions jointly by age and lifetime order. We use instead a calculation of women's mean age at abortion in each of the two states to allocate abortions by their order for the woman, shown below. The age distributions of abortion patients are from state health department reports (for Delaware) (*2016 Annual Report - Delaware Health and Social Services - State of Delaware*, 2016; *2017 Annual Report - Delaware Health and Social Services - State of Delaware*, 2017; *2018 Annual Report - Delaware Health and Social Services - State of Delaware*, 2018; *2019 Annual Report - Delaware Health and Social Services - State of Delaware*, 2019; *2020 Annual Report - Delaware Health and Social Services - State of Delaware*, 2020; *Abortion 2018 Ss6907 - Supplementary Tables 1-17 with 2017 Data*; Jatlaoui et al., 2019; Kortsmit, 2020, 2021) and Guttmacher tabulations using information from surrounding states (for Maryland) (Maddow-Zimet & Kost, 2021). The share of prior abortions (abortion order) are those released in state health department reports and CDC compilation tables. These are available for Delaware but not for Maryland. We assume therefore that the proportion of first versus higher-order abortions for Delaware by year applies equally to annual abortions in Maryland.

We use a simple probability calculation to derive incidence of first abortions, d, in the 2016-2020 period $d(2016\text{-}2020)$ from all abortions in 2016-2020. We first divide all abortions reported into first-order (of her lifetime) abortions, d_1, and higher-order abortions (of her lifetime), d_{2+}. We define by q the probability that a higher-order abortion during the period 2016-2020 was preceded by an abortion that occurred also in the period 2016-2020. Thus, the total incidence of abortions in the period 2016-2020 is given by the incidence of first-order abortions d_1 plus the incidence of higher-order abortions d_{2+} after multiplying them by the proportion $[1-q]$ of higher-order abortions that were the first to occur to the woman in the 2016-2020 period:

$$d(2016\text{-}2020) = d_1(2016\text{-}2020) + d_{2+}(2016\text{-}2020)*[1-q] \qquad (1)$$

The calculation of $[1-q]$ is simplified by our assumption that a woman had at most one abortion each year. Using this assumption, we calculate the probability that a woman who is observed to have an abortion in a year between 2017 and 2020 had an abortion in the year(s) since 2016. Thus, if the observed year of her higher-order abortion is 2017, we need only calculate the probability she had a previous abortion in the year 2016. If the observed year of her higher-order abortion is 2020, we need to calculate the probability she had a previous abortion in any of the years 2019, 2018, 2017, or 2016. We use data on the distribution of abortions by age to do this. Because age at abortion is grouped in five-year categories in our CDC data source, we assume a constant probability that any prior abortion occurred in any of the previous years 2019, 2018, 2017, and 2016. We denote this constant annual probability by q_s. This leads to our re-expressing equation (1) by:

$$d(2016\text{-}2020) = d_1(2016\text{-}2020)$$

$$+ d_{2+}(2016)$$

$$+ d_{2+}(2017)*[1- q_s]$$

$$+ d_{2+}(2018)*[1- q_s]*[1- q_s]$$

$$+ d_{2+}(2019)*[1- q_s]*[1- q_s]*[1- q_s]$$

$$+ d_{2+}(2020)*[1- q_s]*[1- q_s]*[1- q_s]\ ]*[1- q_s]$$

Defining the number of years since 2016 by y, the above equation can then be simplified to:

$$d(2016\text{-}2020) = d_1(2016\text{-}2020) + \sum d_{2+}(y)*[1- q_s]^y \qquad (1^*)$$

In words, the incidence of distinct abortions in the period 2016 to 2020 is the sum of all first abortions (of the woman's lifetime) between 2016 and 2020 and those higher-order abortions between 2016 and 2020 that were not preceded by a previous abortion that occurred in any of the *y* years since 2016.

Table S4: State of Delaware, Distribution of Prior Lifetime Abortions by Year

Calendar Year	2016	2017	2018	2019	2020
Percent 0 Prior Abortions	58.3	63.5	63.5	61.5	61.0
Percent 1 Prior Abortions	24.9	23.0	23.4	23.3	23.2
Percent 2 Prior Abortions	10.0	9.1	8.7	9.7	9.4
Percent 3+ Prior Abortions	6.8	4.3	4.4	5.4	6.4

We assume that the distribution of first and higher-order abortions in Maryland is identical to that of Delaware (Table S4 above), and therefore apply the annual proportion of first abortions from the Delaware data to both the Delaware and Maryland total number of abortions for each year in each of the two states. This gives us the first term of the right hand side of equations (1) and (1*), $d_1(2016\text{-}2020)$. To calculate the second term of the right hand side of equation (1), $d_{2+}(2016\text{-}2020)*[1-q]$, we treat each year of observed abortions separately, in recognition that the later the year in which the higher-order abortion was observed, the greater the likelihood that it was preceded by another abortion since 2016. As shown in the longer form of the equation, we nevertheless apply a constant fraction, q_s, to each prior year. That is, to calculate $d_{2+}(2016\text{-}2020)*[1-q]$, we need first to estimate q_s. We do this separately for Delaware and Maryland by using the distributions of abortions by state, together with two assumptions. The first is that all higher-order abortions occur at the mean age at abortion for that state. The second is that the probabilities of abortion between one year and the previous year are independent. These are assumptions that we are forced to make in the absence of data on the joint distribution of abortions by age, order, and numbers of years between abortions. We use abortion incidence calculated at the mean age of abortion, where annual abortion rates are at their peak, as a crude proxy for the likely actual positive correlation between the probabilities of abortion between one year and the previous year. The use of this proxy device reduces the likelihood of underestimation of repeat abortions and thereby reduces the likelihood of our overestimating abortions to distinct women.

We estimate q_s at .0816 in Delaware and .0853 in Maryland. That is, we estimate that in each year since 2016 and prior to the year of the observed higher-order abortion, the woman had a .0816 probability in Delaware, and a .0853 probability in Maryland, of having had a prior abortion in a given year prior to the observed higher-order abortion. The values of these two probabilities are derived as follows. We assume that in year t = 2016, 2017,…, 2020, in which the higher-order abortion is observed, the abortion occurs at the mean age within the state at which women have abortions. We make this assumption because we know only that the woman has had a higher-order abortion, and we have no information on the ages at which women have higher-order abortions (the CDC distributions are either by age or by abortion order, but not jointly by age and order). We calculate this mean age at a'=27, from data on Delaware women's ages at abortion compiled by the CDC over the years 2016 to 2020. Our calculation of q_s in Maryland follows the same form; however, the state of Maryland does not release statistics on the age distribution of abortion patients nor on the distributions of first and higher order abortions. For Maryland, we use Guttmacher Institute imputed age distributions of abortions, which they derive as an average of age distributions reported by neighboring states (simple average of Delaware, District of Columbia, Pennsylvania, Virginia and West Virginia) for 2016 and 2017 (Maddow-Zimet & Kost, 2021). For the years 2018-2020, we assume the age distribution is constant at its imputed 2017 values. The resulting age distributions differ only slightly between Delaware and Maryland, and the mean age at abortion is the same for the two states at a'=27.

Next, we need an estimate of the probability that a woman observed to have a higher-order abortion in year t mean age at a'=27 would have had a previous abortion in the year t-1, when she was on average age a'-1=26. We calculate this as the proportion $p(26)$ of abortions at the single-year age $a'-1=26$ among all abortions occurring to women at ages below the mean age at abortion. This proportion is then our estimate of q_s. It is calculated by: $q_s = p(26) / \sum_{a=15,..,26} p(a)$. In words, it is the probability that, conditional on having had an abortion before age 27, the abortion occurred exactly one year ago at age 26. We estimate the distribution of abortions by age up to the mean age at abortion using the average age distribution of years 2016-2020 obtained from the CDC compilations (for 2016-2019) and state health department reports (for 2020). Because we know the mean age at abortion only in five-year age groups, and our calculation is that 25-29 is the five-year age group containing the mean age at abortion, we then use a constant value of q_s to represent the probability a woman had an abortion at any of the years since 2016, when she is assumed to have always been in the age grouping of the mean age at abortion. In Table S5, we summarize our full list of data sources and assumptions for each state.

Table S5: Data Sources and Assumptions by State

S5A: State of Delaware

Year	**Parameter**			
	Population of Women Aged 18-44	Age Distribution of Abortion Patients	Number of Abortions Among Women 18-44	Share of Prior Lifetime Abortions
2016	ACS	State Health Department Report Table D13	Calculated by Guttmacher Institute and released in Pregnancies, Births and Abortions in the United States, 1973–2017	CDC Compilation Report, Table 17
2017	ACS	State Health Department Report Table D13	Calculated by Guttmacher Institute and released in Pregnancies, Births and Abortions in the United States, 1973–2017	CDC Compilation Report, Supplemental Table 8
2018	ACS	State Health Department Report Table D13	Calculated by authors as average of 2017 and 2019 values	CDC Compilation Report, Table 8
2019	ACS	State Health Department Report Table D13	Calculated by authors using total abortions by all state residents in 2019-2020 Guttmacher Abortion Provider Census multiplied by the share of all abortions had by women 18-44	CDC Compilation Report, Table 9
2020	ACS	State Health Department Report Table D13	Calculated by authors using total abortions by all state residents in 2019-2020 Guttmacher Abortion Provider Census multiplied by the share of all abortions had by women 18-44	State Health Department Report Table D13

S5B: State of Maryland

Year	Parameter			
	Population of Women Aged 18-44	Age Distribution of Abortion Patients	Number of Abortions Among Women 18-44	Share of Prior Lifetime Abortions
2016	ACS	Calculated by Guttmacher Institute using average of surrounding states and released in Pregnancies, Births and Abortions in the United States, 1973–2017	Calculated by Guttmacher Institute and released in Pregnancies, Births and Abortions in the United States, 1973–2017	Use Delaware values as reported in CDC Compilation Report, Table 17
2017	ACS	Calculated by Guttmacher Institute using average of surrounding states and released in Pregnancies, Births and Abortions in the United States, 1973–2017	Calculated by Guttmacher Institute and released in Pregnancies, Births and Abortions in the United States, 1973–2017	Use Delaware values as reported in CDC Compilation Report, Supplemental Table 8
2018	ACS	Assumed constant from 2017 values.	Calculated by authors as average of 2017 and 2019 values	Use Delaware values as reported in CDC Compilation Report, Table 8
2019	ACS	Assumed constant from 2017 values.	Calculated by authors using total abortions by all state residents in 2019-2020 Guttmacher Abortion Provider Census multiplied by the share of all abortions had by women 18-44	Use Delaware values as reported in CDC Compilation Report, Table 9
2020	ACS	Assumed constant from 2017 values.	Calculated by authors using total abortions by all state residents in 2019-2020 Guttmacher Abortion Provider Census multiplied by the share of all abortions had by women 18-44	Use Delaware values as reported in State Health Department Report Table D13

Additional notes:

i. We use the population of women in the denominator, but not all abortion patients identify as female. Unfortunately, we do not have data on the share of male identified and non-binary abortion patients.

References

2016 Annual Report—Delaware Health and Social Services—State of Delaware. (2016). https://dhss.delaware.gov/dph/hp/2016.html

2017 Annual Report—Delaware Health and Social Services—State of Delaware. (2017). https://dhss.delaware.gov/dph/hp/2017.html

2018 Annual Report—Delaware Health and Social Services—State of Delaware. (2018). https://dhss.delaware.gov/dph/hp/2018.html

2019 Annual Report—Delaware Health and Social Services—State of Delaware. (2019). https://dhss.delaware.gov/dph/hp/2019.html

2020 Annual Report—Delaware Health and Social Services—State of Delaware. (2020). https://dhss.delaware.gov/dph/hp/2020.html

Abortion 2018 ss6907—Supplementary Tables 1-17 with 2017 data. (n.d.). Retrieved July 26, 2022, from https://stacks.cdc.gov/view/cdc/96608

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, *20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Guttmacher Data Center. (n.d.). Retrieved July 26, 2022, from https://data.guttmacher.org/regions

Jatlaoui, T. C., Eckhaus, L., Mandel, M. G., Nguyen, A., Oduyebo, T., Petersen, E., & Whiteman, M. K. (2019). *Abortion surveillance—United States, 2016*. https://www.cdc.gov/mmwr/volumes/68/ss/ss6811a1.htm

Kissling, A., & Jackson, H. M. (2022). Estimating prevalence of abortion using list experiments: Findings from a Survey of Women in Delaware and Maryland. *Women's Health Issues*, *32*(1), Article 1. https://doi.org/10.1016/j.whi.2021.08.003

Kortsmit, K. (2020). Abortion Surveillance—United States, 2018. *MMWR. Surveillance Summaries*, *69*. https://doi.org/10.15585/mmwr.ss6907a1

Kortsmit, K. (2021). Abortion Surveillance—United States, 2019. *MMWR. Surveillance Summaries*, *70*. https://doi.org/10.15585/mmwr.ss7009a1

Lépine, A., Treibich, C., & D'Exelle, B. (2020). Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours. *Social Science & Medicine*, *266*, 113326. https://doi.org/10.1016/j.socscimed.2020.113326

Maddow-Zimet, I., & Kost, K. (2021). *Pregnancies, births and abortions in the United States, 1973–2017: National and state trends by age*.

Ruggles, Steven, Flood, Sarah, Goeken, Ronald, Schouweiler, Megan, & Sobek, Matthew. (2022). *IPUMS USA: Version 12.0* (12.0) [Data set]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V12.0

Wolak, F. A. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica: Journal of the Econometric Society*, 981–995.